

UNITED STATES DISTRICT COURT
WESTERN DISTRICT OF NEW YORK

M.O.C.H.A. SOCIETY, INC., et al.,

Plaintiffs,

-vs-

98-CV-99C

CITY OF BUFFALO, et al.,

Defendants.

APPEARANCES: THOMAS S. GILL, ESQ., Frederick, Maryland, and HARVEY P. SANDERS, ESQ., Cheektowaga, N.Y., for Plaintiffs.

HODGSON RUSS LLP (ADAM W. PERRY, ESQ., JOSEPH S. BROWN, ESQ., and JOSHUA ISAAC FEINSTEIN, ESQ., of Counsel), Buffalo, New York, for Defendant City of Buffalo.

In Second Amended Complaint "B" in this case, plaintiffs Men of Color Helping All ("M.O.C.H.A.") Society, Inc., claim that the City of Buffalo's use of a statewide promotional examination given in 1998 to generate a list for promoting Buffalo firefighters to the rank of lieutenant (the "1998 Lieutenant's Exam") had a disparate impact against African-American firefighters, in violation of Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e-2 (see Item 54). An evidentiary hearing was held by the court over the course of five days in June - August 2008 for the limited purpose of determining whether the 1998 Lieutenant's Exam was "job related for the position in question and consistent with business necessity," as required under Title VII to validate the use of the Exam notwithstanding disparate impact. See 42 U.S.C. § 2000e-2(k)(1)(A)(i).

Based on the testimony and evidence presented at the hearing, and considering the arguments set forth in the parties' post-hearing submissions, the court makes the following findings of fact and conclusions of law.

I. Background

In December 1997, the City of Buffalo sent a request to the New York State Department of Civil Service for services related to the preparation and scoring of an examination for promotion of firefighters to the position of Fire Lieutenant (see "Request for Examination Assistance," Joint Exhibit 9; see also Tr. 363-64).¹ The request was made in accordance with Section 23 of the New York Civil Service Law, which provides:

The state civil service department, upon the request of any . . . municipal commission, shall render service relative to the announcement, review of applications, preparations, construction, and rating of examinations, and establishment and certification of eligible lists for positions in the classified service under the jurisdiction of such municipal commission.

N.Y. Civil Service Law § 23(2).

The Department of Civil Service, Testing Services Division, is the State agency responsible for providing examination assistance to local jurisdictions (Tr. 204-05). The Testing Services Division employee responsible for overseeing the development of the 1998 Lieutenant's Exam was Associate Personnel Examiner Wendy J. Steinberg, Ph.D., under the supervision of Testing Services Director Paul D. Kaiser (Tr. 14, 107-08). The official title of the Exam is the "Lower Level Fire Promotion Series" (Tr. 118). It is used by some jurisdictions for promotion to the rank of Captain (*id.*).

¹Numerical references preceded by "Tr." are to pages of the trial transcript, filed with the Clerk of the Court as separate Docket Items 361, 362, 363, 370, and 377.

The City administered the Exam on March 14, 1998. A total of 179 White firefighters and 89 African-American firefighters took the Exam. Of those who took the Exam, 133 White candidates passed and 46 failed, for a pass rate of 74.3 percent; and 38 African-American candidates passed and 51 failed, for a pass rate of 42.6 percent (see Plaintiffs' Exhibit 25).

M.O.C.H.A. originally filed this action in February 1998 (before the March 1998 administration of the Exam) seeking declaratory relief and damages on their own behalves and as representatives of a proposed class of all African-American firefighters employed by the City during the three prior years, based upon allegations relating to the City's overall policy for promoting firefighters to the rank of lieutenant, as well as the City's drug-testing policy as implemented within the Buffalo Fire Department. Due to the complex nature of the original pleading, the court directed plaintiffs to file separate amended complaints setting forth the claims relating to the drug-testing program (Complaint "A") and the promotional practices (Complaint "B"), and subsequently granted leave to file second amended complaints dealing with each claim.

Following protracted discovery and motion practice, M.O.C.H.A. twice moved for summary judgment on Second Amended Complaint B arguing that the statistical results of the City's administration of the Lieutenant's Exam in March 1998 constituted *prima facie* evidence that the Exam had a disparate impact on African-American firefighters, and that the City failed to meet its burden to demonstrate that the Exam was valid—*i.e.*, that the Exam was job-related and consistent with business necessity. The court denied these motions, finding on each occasion genuine issues of material fact with respect to the validity of the Exam, see *M.O.C.H.A. Society, Inc. v. City of Buffalo*, 2005 WL 589834

(W.D.N.Y. February 28, 2005); *M.O.C.H.A. Society, Inc. v. City of Buffalo*, 2007 WL 3354211 (W.D.N.Y. November 9, 2007), setting the stage for the evidentiary hearing which took place during the summer of 2008.

At the hearing, Dr. Steinberg and Mr. Kaiser testified about the services they performed as the Civil Service Department officials charged with the task of developing the 1998 Lieutenant's Exam. The court also heard testimony from plaintiffs' employment testing expert, Kevin R. Murphy, Ph.D., and from the City's Director of Civil Service, Ms. Olivia Licata.

The hearing testimony is summarized here as follows:

Wendy Steinberg

Dr. Steinberg was called as a witness by both parties to testify about the steps she took to develop the 1998 Lieutenants' Exam. Her educational background includes extensive course work in the development and validation of employment tests, culminating in a Ph.D. in Educational Psychology and Statistics from the State University of New York at Albany (Tr. 97-99). She was employed by the Testing Services Division from 1979 to 1998, and was eventually promoted to the position of Associate Personnel Examiner (Tr. 13-14, 99, 107). During her tenure as Associate Personnel Examiner, she was responsible for supervision and oversight of the development of several employment examinations, including the 1998 Lieutenants' Exam (Tr. 14, 107-08).

Dr. Steinberg's oversight duties included determining the validity of the Exam, relying on the following professional standards (which she referred to as the "joint standards"): the American Psychological Association Standards for Educational and

Psychological Tests (the “APA Standards”) (Defendants’ Exhibit 132); the Principles for the Validation and Use of Personnel Selection Procedures, authored by the Society for Industrial and Organizational Psychology (the “SIOP Principles”) (Defendants’ Exhibit 133); and the Uniform Guidelines on Employee Selection Procedures (the “Uniform Guidelines”) adopted by the Equal Employment Opportunities Commission (“EEOC”) (Tr. 108-11). She also relied on comments from the Fire Advisory Committee, a standing committee of fire personnel that oversees the Civil Service Department’s efforts to develop tests for fire-related jobs (Tr. 122-23). During the time Dr. Steinberg was engaged in the development of the Exam, the Fire Advisory Committee consisted of personnel from the Office of Fire Prevention and Control, the Office of Career Chiefs, and the New York State Professional Firefighter Association. The committee also included one member from a large fire department, one member from a small fire department, and one member representing racial or ethnic diversity review issues (Tr. 120-21). Dr. Steinberg testified that the Fire Advisory Committee’s primary role was to assist in the development of the job analysis for the new examination, summarized in a document she prepared entitled “1995-97 Fire Service Job Analysis” (Joint Ex. 7).

As an initial step in conducting the job analysis, Dr. Steinberg prepared a “task rating list” describing the different duties of fire personnel in all positions, from fire truck driver to fire chief. In developing the task rating list, Dr. Steinberg reviewed the job specifications submitted by Buffalo and various other jurisdictions, along with task rating lists from previous job analyses, and submitted the information to the Fire Advisory Committee for further suggestions on the tasks to be included on the list. The completed list of 190 tasks was sent out as a survey in April 1995 to all full-time paid incumbent fire personnel in all

fire departments in New York State (outside of New York City), asking them to rate such things as the frequency they perform the listed tasks, the importance of those tasks, the amount of time they spend on those tasks, and when they needed to perform those tasks (Tr. 123-24, 126-29).

The next step in the job analysis process was to prepare and distribute to incumbents a survey of the skills, knowledges, abilities, and personal characteristics ("SKAPs") employees in each job title needed to perform the tasks identified in the task survey. The SKAP survey was prepared in a manner similar to the task survey, with review and editing by the Fire Advisory Committee. The final version listing 150 SKAPs was sent out to incumbents in October 1996 (Tr. 129-33; Joint Ex. 15). Dr. Steinberg also sent surveys to 13 other jurisdictions nationwide, including large municipalities such as Los Angeles, Baltimore, Miami, Denver, New York City, the District of Columbia, and Chicago, seeking information regarding the minimum qualifications and test plans they employ for all level of competitive fire positions under their jurisdictions (Tr. 135-38; Joint Ex. 7, p. 12). According to Dr. Steinberg, the test plans used by these jurisdictions were very similar to the 1998 Lieutenants' Exam (Tr. 138).

Dr. Steinberg reported in her job analysis summary that several larger jurisdictions, including Buffalo, refused to participate in the task/SKAP survey at a meaningful level (see Joint Ex. 7, pp. 6, 11). When questioned about this by plaintiffs' counsel, Dr. Steinberg testified that she sent Buffalo directly over 900 surveys to cover 833 fire positions, and followed up with two additional mailings through Civil Service and the statewide union office, but she received only 68 responses to the task survey and no responses to the SKAP survey (Tr. 54-55; 59-62; see also Joint Exs. 13, 17). According to Dr. Steinberg,

this was insufficient data to determine whether tasks needed for the fire lieutenant job in Buffalo required the same knowledge, skills, and abilities as in other jurisdictions, so she compared the responses she received to the responses from other large fire departments in New York State and nationwide, and then compared that data to information about the fire lieutenant job obtained from the Fire Advisory Committee (Tr. 67-68). She also testified that while she was not aware of any studies showing that the fire lieutenant job in Buffalo is so unique that it differs significantly from other large fire departments statewide and nationwide, there are several “validity generalization” studies available to show that tasks for similar positions require the same knowledge, skills, and abilities regardless of where they are performed (Tr. 69-70).

Dr. Steinberg testified that after collecting the responses to the task rating and SKAP surveys, the next step in her job analysis was to analyze the data to determine what topics to include on the Exam. This included determining how many responses had been received, determining the average values for each of the questions asked about each of the tasks, and assigning an overall value to each task. She then consulted with the Fire Advisory Committee to set an appropriate cutoff value for determining the tasks to be included on the Exam, and to determine which tasks should be grouped into similar areas to develop appropriate subtests (Tr. 139-40).

Dr. Steinberg identified two methods for developing the subtests: rational grouping, which involves the use of ordinary judgment to combine similar tasks for inclusion in a subtest, and factor analysis, which is a statistical technique used to group survey data by similarity of response. Dr. Steinberg testified that the subtest groupings developed from the two methods matched “almost perfectly” (Tr. 140).

Dr. Steinberg identified several documents related to her job analysis. Defendants' Exhibit 107 is a copy of the task listing sent out to all incumbents, with Dr. Steinberg's handwritten indication of the overall values assessed for tasks considered critical to performing the job of lieutenant. Defendants' Exhibit 123 is a document entitled "Fire Service Task Criticality Ratings" which Dr. Steinberg described as the guidelines she used to combine the different factors of importance, frequency, and consequence of error into an overall rating for each task. Defendants' Exhibit 108 is a handwritten chart entitled "Task Criticality Ratings" which indicates overall importance, frequency, and consequence of error ratings for each of the 190 tasks and for each job title (Tr. 140-43). Joint Exhibit 21 is a series of computer printout sheets identified in Dr. Steinberg's handwriting as "Factor Analysis for Critical Day 1 Tasks" for the job title of fire lieutenant. Dr. Steinberg testified that she utilized the information on these computer printouts to determine which tasks should be grouped together to develop the questions for the fire-related subtests of the 1998 Lieutenants' Exam (Tr. 145-46).

Defendants' Exhibit 106 is a document entitled "Fire Service Task-SKAP Linkage," which identifies the SKAPs needed to do each of the listed tasks (Tr. 147-48). This document was developed upon consultation with a panel of Testing Services Division personnel and was approved by the Fire Advisory Committee as another way of determining the appropriate subtest areas for the Exam (Tr. 148).

Joint Exhibit 20 is a document entitled "Final Scope," which identified six subtest areas for the 1998 Lieutenants' Exam: (1) fire attack and suppression, (2) fire prevention, (3) rescue and first responder, (4) understanding and interpreting written material, (5) training practices, and (6) supervision. Subtests (1), (2), and (3) are referred to as the

“fire-related” subtests, and subtests (4), (5), and (6) are referred as the “generic” or “cross-occupational” subtests (Tr. 149-50). Dr. Steinberg testified that she included the three generic subtests on the Exam because the ratings of the tasks associated with these activities were very high, and consultation with the Fire Advisory Committee and the Office of Fire Prevention and Control confirmed that these activities were important parts of the fire lieutenant’s job. These subtest areas were also included in the test plans for the title of fire lieutenant from other jurisdictions nationwide, and were identified as knowledges, skills, and abilities (“KSAs”) required for the job of fire lieutenant on the job specifications provided by the statewide municipalities, including Buffalo (Tr. 150-54). The questions for the generic subtests were developed by Civil Service Department units specializing in cross-occupational testing upon request of Dr. Steinberg, with reference to the results of the job analysis, the listings of required tasks and KSAs, and the particular municipality’s job specifications (Tr. 158-60).

Dr. Steinberg also testified that she recruited subject matter experts (“SMEs”) from all jurisdictions across New York State for input to assist in the development of the questions to be included on the fire-related subtests. Defendants’ Exhibit 119 is a copy of the Department of Civil Service memorandum, dated October 21, 1997, that was sent to local civil service agencies statewide requesting nominees and listing the qualifications for SMEs to help develop questions for the fire-related subtests. Dr. Steinberg held a two-day meeting with the SMEs in December 1997 at the Fire Academy in Montour Falls, at which the SMEs were given instructions about how to write test questions and were provided with a list of proposed topics (Defendants’ Ex. 118). The SMEs drafted questions, which were reviewed and edited first by Dr. Steinberg and her supervisor Paul Kaiser, and then by the

Fire Advisory Committee (Tr. 154-58; see *also* Defendants' Exs. 113, 114, and 115).²

Although invited to participate, the Buffalo Fire Department did not send any SMEs to the December 1997 meeting in Montour Falls (Tr. 86-87).

Paul Kaiser

Paul Kaiser has been employed by the New York State Civil Service Department since 1972 (Tr. 180-81). He is currently Director of the Testing Services Division, which provides examination services to all State agencies, and upon request to local jurisdictions, for development and scoring of written examinations to be used in the selection of employees for competitive class positions (Tr. 204-05).

Mr. Kaiser worked his way up through the ranks of the Testing Services Division, and is familiar with all aspects of the examination development process. The entry level position is Personnel Examiner Trainee. Upon satisfactory completion of the traineeship, the employee is eligible for promotion to the position of Senior Personnel Examiner, with primary responsibility for the review of the material to be included on a particular examination. Upon successful completion of an examination, the employee is eligible for promotion to the supervisory position of Associate Personnel Examiner, which is the position held by Dr. Steinberg at the time the 1998 Lieutenant's Exam was developed. The next position is Principal Personnel Examiner, with responsibility for supervision of the Associate Examiners and their units, followed by the positions of Chief Personnel Examiner, Assistant Director, and finally Director (Tr. 205-06).

²Defendants' Exs. 113, 114, and 115 are copies of the original test questions written by the SMEs and submitted to the Fire Advisory Committee for review, editing, deletion, and selection for inclusion on the 1998 Lieutenant's Exam. These exhibits have been received in evidence under seal.

Mr. Kaiser testified that all examinations developed by the Testing Services Division are subject to rigorous internal review, guided by the joint standards (*i.e.*, the APA Standards, the SIOP Principles, and the EEOC's Uniform Guidelines) (Tr. 208). At a minimum, the material to be included in a particular examination is reviewed by either the Senior Personnel Examiner or the Associate Personnel Examiner (in this case, Dr. Steinberg), who would have primary supervisory responsibility for developing the examination. The material is also subjected to a process called "pre-rating review," whereby candidates who have sat for an examination have an opportunity to review and object to test questions prior to final scoring. The examination may be exempted from pre-rating review if it has received prior approval from the Civil Service Commission. Under the prior approval process, the exam questions are reviewed initially by a Senior Personnel Examiner, then by an Associate Personnel Examiner, and then by a section head, who is either a Principal or a Chief Personnel Examiner. The exam material is then submitted for approval by the Assistant Director and, finally, by the Director (Tr. 209-10). The 1998 Lieutenant's Exam received prior approval from the Civil Service Commission after being subjected to these numerous levels of internal review (Tr. 211).

Mr. Kaiser testified that the Testing Services Division is organized by occupational specialty, with certain units designated as cross-occupational to develop examination material which focuses on competencies that are common across numerous occupations rather than on knowledge unique to a particular field (Tr. 211-12). He testified at some length about the Division's routine practices for selecting questions to be included on cross-occupational subtests, and explained the different factors that would be considered by the different subunits responsible for the three subtests which were included on the

1998 Lieutenant's Exam (Tr. 212-19). For example, the supervisory subunit would consider such factors as the level of supervisory responsibilities of the job, and whether the supervision involved a white-collar (or clerical) or blue collar (out in the field) work environment. The subunit would also look at the history of the subtest to determine how well the questions have functioned on previous examinations (Tr. 218-19; see also Defendants' Ex. 135).

Mr. Kaiser testified that the overall process of validating a vocational examination involves the accumulation of evidence which provides information that the test is appropriate for its purpose (Tr. 226). This evidence includes the historical data about past use of the subtests, objections from the jurisdictions requesting the exams, and information obtained from individuals with experience in the jobs for which the exams are being developed (Tr. 227-34).

Kevin Murphy

As reflected in his considerable curriculum vitae (Plaintiffs' Ex. 24), Dr. Murphy has a Ph.D. in Industrial/Organizational Psychology. He has taught, lectured, and published extensively in the areas of employment testing, including job analysis, performance assessment, and validation, and has provided testimony in several employment testing cases over the course of his 25-year career (Tr. 250-51).³

Dr. Murphy testified that he prepared a "Statistical Analysis of the Scores by White and African-American Test Takers on the 1998 Buffalo Fire Lieutenant Examination"

³Defendants have not challenged Dr. Murphy's qualification to testify in this matter as an expert on employment testing.

(Plaintiffs' Exhibit 25), which reported pass rates of 74.3% for White examinees, and 42.6% for African-American examinees (Tr. 252-53). This resulted in an "adverse impact ratio" of .573, significantly lower than the "four-fifths" or "eighty percent" rule adopted by the EEOC's Uniform Guidelines, which allows an inference of adverse impact to be drawn where the passing rate for a disadvantaged group is less than 80 percent of the passing grade of the highest scoring group (Tr. 253-54).

Dr. Murphy also performed a statistical analysis of the scores on the Exam as a whole and on each of the separate subtests of the Exam, using standardized formulas for eliminating random chance or sampling error (Tr. 254-55). According to Dr. Murphy, this analysis revealed significant systematic differences in favor of White applicants, indicating that the Exam "had a very substantial effect on the employment opportunities of black versus white examinees." (Tr. 255).

In addition to conducting a statistical analysis of the test scores, Dr. Murphy's duties as plaintiffs' employment testing expert included evaluating the validity of the 1998 Lieutenant's Exam. He testified that he performed his evaluation in accordance with the joint standards—including the APA Standards and the SIOP Principles, which he helped develop—and concluded that there was insufficient credible evidence put forth to demonstrate that the Exam was job-related or consistent with business necessity (Tr. 256-57).

Dr. Murphy testified that in his opinion, the Civil Service Department's attempt to use a content-related strategy to validate the three fire-related subtests was deficient in several respects. First and foremost, Dr. Murphy was of the opinion that Dr. Steinberg did not conduct a suitable job analysis. He gave the explanation that, while the original test plan

to conduct a task and SKAP survey of fire departments of all types and sizes was a very sound and reasonable one, the discrepancies in the response rates to the survey—particularly, the low response rates from larger fire departments, and the City of Buffalo's failure to respond in a meaningful fashion at all—resulted in insufficient data to determine whether the content of the Exam reflected the content of the fire lieutenant job in Buffalo. In Dr. Murphy's view, because the job of fire lieutenant is unlikely to be the same in places that are quite different in terms of structure, size, or complexity of their fire departments, the job analysis should have included the collection of data that would allow the test preparers to determine whether the job is, in fact, the same or similar from one jurisdiction to the next. According to Dr. Murphy, when the testing jurisdiction is relying entirely on content validity for a test that has a demonstrated adverse impact against a particular group of examinees, there is a strong responsibility to show that the test measures the same thing that the job entails. Here, the job analysis conducted by Dr. Steinberg was based primarily upon the assumption that the job is similar from one place to another, without any detailed analytic demonstration to support the assumption (Tr. 258-64).

Dr. Murphy also criticized Dr. Steinberg's reliance on validity generalization studies to infer that the content of the fire lieutenant job in Buffalo is the same as in other jurisdictions. According to Dr. Murphy, validity generalization is a method widely used in criterion-related test validation studies, but has little relevance in content-related validation studies. Validity generalization is based upon the notion that criterion-based statistical evidence demonstrating a correlation between test scores and subsequent measures of job performance in one jurisdiction can be used to raise the inference that the same test

will be valid if used in another jurisdiction. Dr. Murphy was not aware of validity generalization ever being used to infer that a job analysis for one position in one location is going to be similar to a job analysis for another person in another location (Tr. 264-67).

Dr. Murphy testified that, because of these inadequacies in Dr. Steinberg's job analysis, there was no way of knowing whether Civil Service used reasonable competence in developing the Exam, whether the content of the Exam related to or was representative of the content of the Buffalo Fire Lieutenant job, or whether the scoring system usefully selected from among the applicants those who could better perform the job (Tr. 268-69). He also testified that there were no validity studies done on the three cross-occupational subtests, nor was there any other evidence in the record to show that any of these subtests measures what it purports to measure or is a valid predictor of job performance (Tr. 269). According to Dr. Murphy, once Dr. Steinberg became aware that the Exam had an adverse impact against Black examinees, it became her absolute responsibility to demonstrate with empirical evidence that the Exam measures what it purports to measure, or that it actually predicts job performance, and she failed to do so (Tr. 273-74).

Based on this analysis, Dr. Murphy concluded that the use of the 1998 Lieutenant's Exam had a substantial adverse impact on the opportunities of Blacks to advance to the position of fire lieutenant, and that the City failed to adequately demonstrate the content validity of the fire-related subtests or to present any evidence whatsoever of any sort of validity study for the cross-occupational subtests (Tr. 279).

Dr. Steinberg's Rebuttal

Over plaintiffs' objection,⁴ the City recalled Dr. Steinberg to rebut several aspects of Dr. Murphy's testimony, particularly his assertion that the low rate of task/SKAP survey responses from Buffalo and other large fire departments resulted in insufficient data to determine whether the content of the Exam reflected the content of the Buffalo fire lieutenant job. Dr. Steinberg testified that the responses she received from 12 largest fire departments surveyed—those having 140 to 150 staff members and at least 10 station houses—far exceeded the number of responses necessary to achieve a 95 percent statistical confidence rate (Tr. 314-15). She also reviewed test plans and other information obtained from several large urban fire departments nationwide, and consulted with the Fire Advisory Committee (Tr. 315-17). She testified that she did not simply assume that fire lieutenants in different jurisdictions perform the same duties; but rather, she relied upon the data in the responses to the fire task listing as compared to data for fire lieutenants nationwide, and as reviewed by the Fire Advisory Committee (Tr. 317). She also performed a discriminate statistical analysis of the fire lieutenant title which showed nearly a 90 percent correlation of duties associated with lieutenant positions in different jurisdictions statewide (Tr. 318-22; Joint Ex. 13).

⁴Prior to Dr. Steinberg's rebuttal, plaintiffs filed a motion *in limine* (Item 373) seeking to preclude her from offering any expert testimony within the scope of Fed. R. Evid. 702 and the disclosure requirements of Fed. R. Civ. P. 26(a)(2). The court denied this motion without prejudice to renew at the close of proof (Item 381). Plaintiffs did renew the motion, and also sought permission to recall Dr. Murphy to rebut Dr. Steinberg's "stealth" expert testimony (see Item 380). The court ultimately denied plaintiffs' motion, finding that Dr. Steinberg's testimony in the case was not subject to the expert witness disclosure requirements because it was based upon personal knowledge obtained as the result of her involvement in the events at issue (Item 393).

Olivia Licata

The final witness called to testify at the hearing was Olivia Licata, current Director of Civil Service Division for the City of Buffalo's Human Resources Department. Her duties as Director of Civil Service include oversight of employment classification, exam qualification and administration, hiring, and payroll certification. She has been with the Civil Service Department since 1974, and has served as Director since 2004 (Tr. 360-61).

In 1997, Ms. Licata held the title of Personnel Specialist I, with oversight responsibility for the Examination Division for the Department. Her duties in that regard included the planning and administration of employment exams, posting exam announcements, and reviewing exam applications. At that time, the City had traditionally relied on the New York State Department of Civil Service to prepare examinations for almost all job titles. Ms. Licata was the officer who signed the Request for Examination Assistance form, dated December 11, 1997 (Joint Ex. 9), requesting examination services for the 1998 Lieutenant's Exam (Tr. 361-63).

The initial step in the process for requesting examination assistance from the State was to review the job specification for the position to be tested. The job specification is the classification for the position, describing the duties, the typical work activities, the KSAs (knowledge, skills, and abilities) required for the position, and the minimum qualifications for the job. The specification would first be reviewed internally by Civil Service, and then sent to the respective City department and eventually to the union for review, comments, and updates. The specification for the fire lieutenant position in Buffalo was initially developed in 1960, and was reviewed and revised several times over the years prior to the development of a new examination or as the need otherwise arose. It was last reviewed

on December 10, 1997—immediately prior to transmittal of the request for the 1998 Lieutenant's Exam (Tr. 362-64).⁵

Ms. Licata testified that the next step in the process involved receipt of a document from State Civil Service entitled "Transmittal of Instructions for Preparation of Announcements," which was dated December 23, 1997 (see Plaintiffs' Ex. 1, at Bates 000016-19). This document contained the examination date, the final filing date, and a description of the subtests to be included in examination announcements to be sent out by the testing municipalities. Upon receiving this transmittal, Ms. Licata compared the

⁵The job specification attached to the Request for Examination Assistance contains the following description of the Buffalo fire lieutenant position:

DISTINGUISHING FEATURES OF THE CLASS

This is a first line supervisory position where incumbents are responsible for the activities of a fire company during an assigned shift. Responsibilities include directing the work of Firefighters at fires and in fire stations, evaluating their work performance and instructing them in new approved firefighting methods. Work is performed in accordance with established procedures and policies as outlined by the Fire Department. The class of Fire Lieutenant is distinguished from that of Fire Captain in that the latter is in charge of the activities of the fire company on all shifts and is in charge of operations at the scene of a fire in the absence of or pending arrival of a superior officer. All work is performed under general departmental regulations and incumbents directly supervise Firefighters under their command.

TYPICAL WORK ACTIVITIES

Responds to all alarms assigned to his company while on duty;
Directs the work of firefighters at scenes of fire and in station house;
Assigns firefighters to lay out and connect hose lines and nozzles, direct hose streams, raise ladders and ventilate buildings;
Inspects property at scene of fire to prevent re-ignition;
Supervises the cleaning, checking and replacement of tools and equipment after a fire;
Inspects personnel, station house, buildings, grounds and facilities to ensure conformity with departmental rules and regulations;
Examines fire trucks and equipment such as ladders and hose to ensure proper order and condition;
Inspects buildings and premises for fire hazards;
Personally supervises a wide variety of cleaning and maintenance tasks performed at the station;
Maintains discipline;
Makes periodic reports of personnel and activities;
performs related duties as required.

(Joint Ex. 9, p. 2).

proposed subtest topics to the information in the fire lieutenant job specification, and determined that the subtests were appropriate for the position. She also sent the transmittal to the Buffalo Fire Department for their input, but her files indicate that she did not receive a response (Tr. 364-66).

Ms. Licata testified that the City no longer uses New York State Civil Service to develop exams for the fire lieutenant position. Instead, in December 2006, the City issued a Request For Proposals ("RFP") to select an outside consulting firm to develop new promotional exams for the Buffalo Fire Department (Defendants' Ex. 130), and in May 2007 entered into an agreement with BE Jacobs, LLC toward that purpose (see Defendants' Ex. 131; Tr. 367-72).

II. Analysis

Title VII provides:

An unlawful employment practice based on disparate impact is established under this subchapter only if—

a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity

42 U.S.C. § 2000e-2(k)(1)(A)(i).

The general standards for establishing unlawful disparate impact are well established. Disparate impact claims involve three stages of proof. First, the plaintiff must make a *prima facie* showing that the employer “uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin”

42 U.S.C. § 2000e-2(k)(1)(A)(i); see *Robinson v. Metro-North Commuter R.R. Co.*, 267 F.3d 147, 160 (2d Cir. 2001). If the plaintiff succeeds, the burden of persuasion then shifts to the employer to demonstrate that the challenged employment practice is valid, *i.e.*, “job related for the position in question and consistent with business necessity.” 42 U.S.C. § 2000e-2(k)(1)(A)(i); see *Gulino v. New York State Educ. Dept.*, 460 F.3d 361, 382 (2d Cir. 2006). Finally, should the employer meet its burden of proving validity, Title VII provides that the employee can still prevail at the third stage by showing that there was an available, equally valid, less discriminatory method of promotion that the employer refused to use. 42 U.S.C. § 2000e-2(k)(1)(A)(ii), (C); see *Robinson*, 267 F.3d at 161.

A. Disparate Impact

As an initial matter, notwithstanding the opportunities presented by the multitude of dispositive motions filed in this case, this court has not definitively ruled that the 1998 Lieutenant’s Exam had a disparate impact on African-American promotional candidates. To make a *prima facie* showing of disparate impact, plaintiffs must (1) identify an employment policy or practice, (2) demonstrate that a disparity exists, and (3) establish a causal relationship between the two. *Robinson*, 267 F.3d at 160. “[S]tatistical proof almost always occupies center stage in a *prima facie* showing of a disparate impact claim . . .,” *id.*, but the statistical disparity must be “sufficiently substantial” to raise the inference that it was caused by the employment practice. *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 994 (1988).

When called upon to make this assessment, the courts have often looked to the Uniform Guidelines’ “four-fifths” rule, referenced above, “under which adverse impact will

not ordinarily be inferred unless the members of a particular race, sex, or ethnic group are selected at a rate that is less than four-fifths of the rate at which the group with the highest rate is selected.” *Watson*, 487 U.S. at 995 n.3 (citing 29 C.F.R. § 1607.4(D)). Though criticized on various technical grounds by the commentators and the courts, see *id.* (citing cases and articles), this standard has provided useful guidance for determining the significance or substantiality of numerical disparities on a case-by-case basis. See, e.g., *Ricci v. DeStefano*, 530 F.3d 88, 112 (2d Cir. 2008); *Green v. Town of Hamden*, 73 F. Supp. 2d 192, 197-98 (D.Conn. 1999); see also *Guardians Association of the New York City Police Department, Inc. v. Civil Service Commission of the City of New York*, 630 F.2d 79, 87-88 (2d Cir. 1980) (entry level exam resulting in minority candidate passing rate of two-fifths that of White candidates had disparate impact “[b]y any reasonable measure, including the . . . four-fifths rule of the EEOC Guidelines”), *cert. denied*, 452 U.S. 940 (1981).

In this case, plaintiffs primarily rely on Dr. Murphy’s testimony, as well as his statistical analysis of the scores for the 1998 Lieutenants’ Exam, which reflect a passing rate for African-American examinees at 57.3 percent of the passing rate for White examinees—significantly less than the Uniform Guidelines’ four-fifths (or 80 percent) standard (see Tr. 253-54; Plaintiff’s Ex. 25). Dr. Murphy’s analysis also indicates that this statistical disparity is not attributable to random chance, sampling error, or any other extraneous factors other than systematic differences in favor of White examinees (see Tr. 254-55).

The City has not forcefully argued to the contrary, nor has it offered proof sufficient to overcome the strong inference of disparate impact raised by Dr. Murphy's testimony and statistical analyses. In her rebuttal testimony, Dr. Steinberg did raise the contention that Dr. Murphy's analyses were flawed in certain ways (see, e.g., Tr. 311-14). However, the City has not offered any countervailing analyses or other documentary evidence to explain or support Dr. Steinberg's criticisms.

Upon review of the record presented at the hearing, and in the absence of any convincing evidence or argument to the contrary, the court has little difficulty finding that plaintiffs have demonstrated a sufficiently substantial statistical disparity in the test scores to raise the inference that the 1998 Lieutenants' Exam had a disparate impact on African-American promotional candidates. Plaintiffs have therefore established their *prima facie* case, and the burden of persuasion has shifted to the City to show that the Exam was sufficiently job-related and consistent with business necessity to overcome the inference of disparate impact.

B. Validity

To demonstrate that a promotional examination is valid notwithstanding its disparate impact, the courts have required the employer to show, by professionally acceptable methods, that the examination is "predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated." *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 431 (1975) (quoting 29 C.F.R. § 1607.4(c)). "The touchstone is business necessity. If an employment practice which operates to exclude [minorities] cannot be shown to be related

to job performance, the practice is prohibited." *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971).

In the Second Circuit, the holding in *Guardians* has long provided the benchmark for courts charged with assessing the validity of an employment examination. Here again, the Uniform Guidelines have proven useful, but *Guardians* recommends a "cautionary approach" to their rigid application. *Guardians*, 630 F.2d at 92. "[T]he Guidelines are not administrative regulations promulgated pursuant to formal procedures established by Congress. They are entitled to deference, not obedience." *Id.* at 91 (internal quotation marks and citations omitted).

According to *Guardians*, the threshold task in determining validity is to select the appropriate method for assessing job-relatedness from the three basic methods recommended by the Uniform Guidelines: criterion-related validation, content validation, and construct validation. *Id.* (citing 29 C.F.R. §§ 1607. 5(B), 1607.14). As described generally in the Guidelines:

[A] criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance. . . . [A] content validity study should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated. . . . [A] construct validity study should consist of data showing that the procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important in successful performance in the job for which the candidates are to be evaluated.

29 C.F.R. § 1607. 5(B)

In this case, Dr. Steinberg's job analysis was conducted along the lines of a content-based study, which traditionally has been considered to be an appropriate validation

method where the test attempts to measure a knowledge or ability, and not a general trait such as intelligence, and the test does not measure a knowledge or ability that an employee is expected to learn on the job. *Guardians*, 630 F.2d at 92; 29 C.F.R. § 1607.14(C)(1). Under *Guardians*, content validation is generally acceptable “as long as the abilities that the test attempts to measure are no more abstract than necessary, that is, as long as they are the most observable abilities of significance to the particular job in question” *Guardians*, 630 F.2d at 93.

Dr. Steinberg’s efforts to develop exam questions measuring important aspects of the fire lieutenant job are clearly reflected in her job analysis (Joint Ex. 7) and incumbent surveys reporting on data obtained from firefighters throughout New York State about the knowledge, skills, and abilities required to perform the tasks of significance to the various positions surveyed (Joint Exs. 11, 15; Defendants’ Exs. 107, 110, 124). The data and proposed questions for the Lower Level Fire Promotion Series were reviewed by a panel of experienced firefighters (see Defendants’ Exs. 120-122), and were compared to testing materials and data for fire lieutenants from other jurisdictions nationwide (see Joint Ex. 7, at p. 12).

In addition, Dr. Steinberg explained that a criterion-related or construct validity study was not feasible in this case because there was no data obtainable by which exam results could be correlated with subsequent job performance. In fact, according to Dr. Steinberg, employment tests are rarely validated using criterion-related studies because reliable correlation of test results and job performance requires data from employees across the entire range of test scores, while civil service systems ordinarily hire from the top end of the test score range. This is one of the reasons the APA Standards have moved away

from a rigid distinction between content, construct, and criterion-related validation, adopting instead “a unitary concept” of validation which assesses “the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose.” Defendants’ Ex. 139, at p. 11; see also Tr. 331-33.

Based on the testimony and evidence presented at the hearing, and according due deference to the recommendations of the Uniform Guidelines as outlined in *Guardians*, the court concludes that Dr. Steinberg’s validation approach was an appropriate method for assessing the job-relatedness and business necessity of the questions to be included on the 1998 Lieutenant’s Exam. Clearly, the three fire-related subtest areas (fire attack and suppression, fire prevention, and rescue/first responder) were designed to measure knowledge, skills, and abilities representative of important aspects of performing the job of fire lieutenant, as opposed to those which a fire lieutenant could be expected to learn on the job, or general traits “such as intelligence, aptitude, personality, commonsense, judgment, leadership, and spatial ability.” 29 C.F.R. § 1607.14(C)(1). The hearing record also supports the conclusion that Dr. Steinberg exercised appropriate professional judgment in relying upon the considerable experience and expertise of the Testing Services Division subunits for development of the generic subtest areas (understanding and interpreting written material, training practices, and supervision), as reviewed by the Fire Advisory Council and cross-referenced against the test plans from other jurisdictions nationwide.

Where content validation has been selected as the appropriate technique for assessing the job-relatedness of an exam, *Guardians* has distilled from the Uniform

Guidelines the following five attributes which the exam should display to be considered valid notwithstanding its disparate racial impact:

- (1) the test-makers must have conducted a suitable job analysis;
- (2) the test-makers must have used reasonable competence in constructing the test itself;
- (3) the content of the test must be related to the content of the job;
- (4) the content of the test must be representative of the content of the job; and
- (5) the test must be used with a scoring system that usefully selects from among the applicants those who can better perform the job.

Guardians, 630 F.2d at 95. The first two attributes "concern the quality of the test's development," while the next three "are more in the nature of standards that the test, as produced and used, must be shown to have met." *Id.* The "essence of content validation" is requirement number three: that the content of the test be related to the content of the job. *Id.*

Each of these criteria is now examined in turn.

1. Job Analysis

The Uniform Guidelines provide the following "technical standard" for conducting a job analysis in a content validity study:

There should be a job analysis which includes an analysis of the important work behavior(s) required for successful performance and their relative importance and, if the behavior results in work product(s), an analysis of the work product(s). Any job analysis should focus on the work behavior(s) and the tasks associated with them. If work behavior(s) are not observable, the job analysis should identify and analyze those aspects of the behavior(s) that can be observed and the observed work products. The work behavior(s) selected for measurement should be critical work behavior(s) and/or important work behavior(s) constituting most of the job.

29 C.F.R. § 1607.14(C)(2).

As reflected in her testimony and summary report (Joint Ex. 7), Dr. Steinberg conducted a comprehensive job analysis designed to update the test plans for all fire service job titles, including fire lieutenant. Her work on this project began in 1994, and continued through 1997 (*id.* at p. 3). A major component of the job analysis was the task/SKAP survey of fire service personnel throughout New York State aimed at gathering detailed job information from incumbents across all titles. The initial task listing survey requested incumbents to provide data regarding whether each of 190 separate tasks was performed on the job, either personally or under the incumbent's supervision; the difficulty, time, frequency, and consequence of error for the task; and whether the task was needed at job entry (see Joint Ex. 11; Defendants' Ex. 107). The second survey sought data from incumbents concerning specific skills, knowledges, abilities, and personal traits associated with each of the critical tasks performed by each job title (Joint Ex. 6; Defendants' Ex. 124). This survey contained 150 items, and requested information to assess the importance of each item to effective job performance as well as the level competence needed at the time of appointment (Defendants' Ex. 109). In preparing both surveys, Dr. Steinberg relied on the job specifications provided by the municipalities, previous job analyses, test plans from other jurisdictions, and input from the Fire Advisory Committee.

Upon receipt of the survey responses, Dr. Steinberg analyzed the data to determine what topics to include on the Exam. This process included linking tasks and SKAPs, rational grouping and factor analysis of survey data to develop subtest groupings, and obtaining input from the Fire Advisory Committee and SME panel. These procedures were

all painstakingly documented (see Joint Exhibit 21; Defendants' Exs. 106, 120-125) and incorporated into the job analysis report.

Plaintiffs' primary challenge to Dr. Steinberg's job analysis is based on Dr. Murphy's criticism that the task/SKAP survey failed to provide sufficient evidence to determine the content of the fire lieutenant position as performed in Buffalo and other large fire departments in New York State. With respect to Buffalo alone, at the time of the survey there were 833 positions in the Buffalo Fire Department, and Dr. Steinberg sent Buffalo approximately over 900 surveys, but even after two follow-up mailings she received only 68 responses to the task survey and no responses to the SKAP survey. According to plaintiffs, this demonstrates that Dr. Steinberg proceeded with her job analysis without sufficient data to determine whether the test plan for the Exam encompassed the knowledge, skills, and abilities required for successful performance of the tasks associated with the job of Buffalo fire lieutenant.

However, despite the failure of Buffalo incumbents to participate in the task/SKAP survey in a manner that could in any way be considered representative, the court's review of the hearing record indicates that Dr. Steinberg nonetheless accumulated substantial evidence to support her job analysis as it applied to the position of fire lieutenant statewide, as well as in the City of Buffalo. First of all, Dr. Steinberg testified that her starting point for developing the task and SKAP surveys was to review the job specifications for the various titles, including the specification for the job of Buffalo fire lieutenant which, as indicated above, contains a detailed description of the duties, the typical work activities, the knowledge, skills, and abilities, and the minimum qualifications required for the position (see Joint Ex. 9). She compared this information to job specifications obtained from other

jurisdictions, including areas outside New York State. Indeed, the court's comparison of the job specifications attached to the requests for examination assistance for the 1998 Lieutenant's Exam submitted by the cities of White Plains, N.Y., and Newburgh, N.Y. (Defendants' Ex. 127) reveals that each contains substantially the same information as Buffalo's job specification with regard to the distinguishing features of the class, typical work activities, and the knowledge, skills, and abilities required for performance of the job of fire lieutenant.

In addition, Dr. Steinberg testified that she received sufficient responses from the twelve largest fire departments in the State to trust the survey results at a 95 percent statistical confidence rate (see Joint Ex. 13), and she also performed a discriminant statistical analysis of the fire lieutenant title which showed approximately a 90 percent correlation of duties associated with lieutenant positions in different jurisdictions statewide (*id.*). Finally, she reviewed and compared test plans and other job information obtained from fourteen large urban fire departments in other parts of the country (see Joint Ex. 7, at p. 12).

Plaintiffs contend that Dr. Steinberg's reliance on information obtained from other jurisdictions is a misapplication of the test development technique known as "validity generalization" which, as Dr. Murphy testified, involves the use of previously developed criterion-related validation studies to infer the validity of the same test used later in another jurisdiction. However, as indicated in her testimony, Dr. Steinberg did not specifically rely on the technique of validity generalization to conduct her job analysis as it related to the 1998 Lieutenant's Exam. Rather, her reference to test plans, job specifications, and other information obtained from outside jurisdictions is consistent with the recognized practice

of accumulating available evidence on the use and validation of similar employment examinations in similar situations from other settings, as outlined in the pertinent APA Standards.⁶

Based on this review, the court finds that Dr. Steinberg's job analysis adequately assesses the important work behaviors required for successful performance of the fire lieutenant job in the City of Buffalo to meet the recommendations of the Uniform Guidelines, as outlined in *Guardians*.

2. Reasonable Competence

As evidenced by the testimony of both Dr. Steinberg and Mr. Kaiser, the Testing Services Division is charged with responsibility for developing employment and promotional examinations for competitive civil service positions statewide, in accordance with New York State Civil Service Law. Testing Services has approximately 100 employees, many of

⁶For example, in discussing the background of test validation procedures, the APA Standards provide as follows:

Gathering [validation] evidence may sometimes involve not only the present instrument in the present situation, but also the available evidence on the use of the same or similar instruments in similar situations. This process involves generalization, either on the basis of common elements, as in synthetic validation, or on the basis of overall job similarity, as in validity generalization.

APA Standards (1985 ed.), Defendants' Ex. 132, at 9.

Further:

When propositions have been identified that would support the proposed interpretation of test scores, validation can proceed by developing empirical evidence, examining relevant literature, and/or conducting logical analyses to evaluate each of these propositions. Empirical evidence may include both local evidence, produced within the contexts where the test will be used, and evidence from similar testing applications in other settings. Use of existing evidence from similar tests and contexts can enhance the quality of the validity argument, especially when current data are limited.

APA Standards (1999 ed.), Defendants' Ex. 139, at 10-11.

whom are professionals with expertise in test development. Dr. Steinberg was employed in the Division for many years, and both she and Mr. Kaiser, her supervisor, have been involved in the development and validation of hundreds of employment tests, including tests for qualification and promotion across all fire service job titles.

In addition, as already discussed, the hearing testimony and evidence amply demonstrate that the Lieutenant's Exam was constructed on the basis of a methodical selection of the tasks critical to job performance, and the knowledge, skills, and abilities needed to perform those tasks. The accumulated data was thoroughly organized, analyzed, rated, reviewed by a panel of experts drawn from all disciplines in the fire service, and compared to similar information gathered from other jurisdictions. This evidence was made available to the subject matter experts who provided assistance in developing test questions for the fire-related subtest areas, as well as to the professionals in the Testing Service Division subunits responsible for developing the cross-occupational subtests.

Based on this showing, the court has little difficulty concluding that the test makers demonstrated reasonable competence in constructing the 1998 Lieutenant's Exam.

3. Job-Relatedness

As discussed above, the job analysis conducted by Dr. Steinberg adequately demonstrates that the content of the 1998 Lieutenants' Exam is related to the content of the job of fire lieutenant in the City of Buffalo. To reiterate, the task and SKAP surveys incorporated detailed job specification information for the position of fire lieutenant obtained from Buffalo and other representative municipalities both inside and outside of

New York State. While the survey responses were inadequate to develop a representative analysis of any direct relationship between the content of the Exam and the content of the job as performed in Buffalo, Dr. Steinberg explained that the responses she received from the twelve largest fire departments in the State were sufficient to achieve a reasonable rate of confidence in the survey results. In addition, Dr. Steinberg's discriminate statistical analysis of the duties associated with the fire lieutenant title statewide showed a substantial correlation between lieutenant positions in different jurisdictions, which was confirmed by her review and comparison of test plans and other job information obtained from several large urban fire departments in other parts of the country.

Based on this showing, the court finds that the job analysis in this case "provides adequate assurance that the identified tasks are in fact the tasks that a [fire lieutenant] performs . . ." *Guardians*, 630 F.2d at 98. Accordingly, the City has demonstrated a sufficient relationship between the content of the 1998 Lieutenants' Exam and the content of the job of fire lieutenant to meet the recommendations of the Uniform Guidelines, as outlined in *Guardians*.

4. Content Representativeness

The Uniform Guidelines also provide that, to demonstrate content validity, the test developer "should show that the behavior(s) demonstrated in the selection procedure are a representative sample of the behavior(s) of the job in question or that the selection procedure provides a representative sample of the work product of the job." 29 C.F.R. § 1607.14(C)(4). Recognizing that strict application of this standard could present both theoretical and practical difficulties, *Guardians* interpreted this standard to require a

showing “that the test measure important aspects of the job, at least those for which appropriate measurement is feasible, but not that it measure all aspects, regardless of significance, in their exact proportions.” *Guardians*, 630 F.2d at 99.

In this regard, the testimony and evidence presented in this case demonstrate that the 1998 Lieutenant’s Exam was developed as an adequate measure of important aspects of the fire lieutenant job as performed in statewide jurisdictions, including Buffalo. As discussed above, Dr. Steinberg’s efforts to construct test questions for the Exam are reflected in her job analysis, reporting on the various methods employed by Testing Services to analyze accumulated survey data about the knowledge, skills, and abilities required to perform tasks of significance to the fire lieutenant job. Among other things, Dr. Steinberg developed a method to rate the criticality of a task according to its importance, frequency of performance, and consequence of error; consulted with the Fire Advisory Committee to set an appropriate cutoff value for determining the tasks to be included on the Exam; and conducted a statistical analysis to determine which of the tasks should be grouped together to develop the questions for the fire-related subtests. She also utilized the task ratings and statistical analysis, along with job specifications and other information about the fire lieutenant position obtained from representative state- and nationwide jurisdictions, in working with the Fire Advisory Committee, subject matter experts, and cross-occupational experts within the Testing Service Division to scope and construct subtest areas representative of the content of the fire lieutenant job.

Based on this showing, the court finds that the 1998 Fire Lieutenant’s Exam was developed as an adequate measure of important aspects of the job of fire lieutenant,

meeting the representativeness requirement of the Uniform Guidelines as interpreted by the Second Circuit in *Guardians*.

5. Scoring System

The final attribute to be examined is whether the City has sufficiently demonstrated that the Exam was used with a scoring system that usefully selected from among the applicants those who can better perform the job. Dr. Steinberg testified in this regard that she determined the passing point for the 1998 Lieutenant's Exam by first establishing the upper limit to the scoring range at 70 percent of the number of questions on the Exam (i.e., 73 correct out of 105 questions), and the lower limit at 60 percent (i.e., 63 correct), in accordance with State Civil Service requirements.⁷ She then set the passing point at 66 correct answers, which resulted in a pass rate of 63 percent (181 out of 287) for Buffalo examinees (see Joint Ex. 23). Statewide, the pass rate for the Exam was 76 percent (571 out of 756) (see Defendants' Ex. 117).

Under the Uniform Guidelines:

Where cutoff scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force. Where applicants are ranked on the basis of properly validated selection procedures and those applicants scoring below a higher cutoff score than appropriate in light of such expectations

⁷Department of Civil Service regulations provide:

The final score of each candidate in a written or oral or other test or combination of tests comprising an examination shall be reported on a scale to 100, where the score of 100 shall represent the best performance possible, expected or attained, and where the score of 70 shall represent a performance meeting the minimum needs of the position to be filled with regard to number and quality of eligibles, and where the score of 60- shall represent any score below 60.

4 N.Y.C.R.R. § 67.1(a).

have little or no chance of being selected for employment, the higher cutoff score may be appropriate, but the degree of adverse impact should be considered.

29 C.F.R. § 1607.5(H). According to the Second Circuit:

This . . . makes sense. No matter how valid the exam, it is the cutoff score that ultimately determines whether a person passes or fails. A cutoff score unrelated to job performance may well lead to the rejection of applicants who were fully capable of performing the job. When a cutoff score unrelated to job performance produces disparate racial results, Title VII is violated. Consequently, there should generally be some independent basis for choosing the cutoff. . . . [A] criterion-related study is not necessarily required; the employer might establish a valid cutoff score by using a professional estimate of the requisite ability levels, or, at the very least, by analyzing the test results to locate a logical “break-point” in the distribution of scores.

Guardians, 630 F.2d at 105.

The court's review of the record in this case indicates that the City has satisfied these requirements. As discussed above, the testimony and evidence presented at the hearing establishes that Dr. Steinberg conducted a suitable job analysis, and used reasonable competence in constructing a test with content related to, and representative of, the content of the job. It is also evident that Dr. Steinberg considered the degree of adverse impact in her analysis of the test scores to establish the passing point for the Exam. At least with respect to the results for the City of Buffalo, setting the passing score at 66 correct answers rather than at the maximum allowable score of 73 resulted in doubling the pass rate for African-American examinees from 19.5 percent (17 of 87) to 40 percent (35 of 87) (see Joint Ex. 23). This provides a reasonable basis for concluding that Dr. Steinberg used professional judgment in analyzing the test results to locate a logical passing point, and that the scoring system for the Exam provided a useful method of selecting those examinees better capable of performing the job of fire lieutenant.

To summarize, the court finds the proof presented at the hearing sufficient to show that the 1998 Lieutenant's Exam was developed by the Testing Services Division of the New York State Civil Service Department in a manner that is significantly correlated with important elements of work behavior which are relevant to the position of fire lieutenant as performed in the City of Buffalo. Accordingly, the City has met its burden of demonstrating that the Exam is job-related for the position and consistent with business necessity, and the court turns its focus to the third stage of proof of disparate impact under Title VII.

C. Alternative Employment Practice

As mentioned above, at the third stage, the burden of persuasion shifts back to the plaintiffs to show that other tests or devices were available for selection of fire lieutenant candidates "that would also satisfy the asserted business necessity, but would do so without producing the disparate effect." *Robinson*, 267 F.3d at 161. Upon review of the record, and despite the long history of this litigation, the court finds no proof that could be considered sufficient to meet this burden of persuasion, or to raise a genuine issue of fact requiring further proceedings in this regard.

CONCLUSION

Based on the foregoing, plaintiffs' Second Amended Complaint "B" is dismissed to the extent it seeks relief under Title VII based on the City of Buffalo's use of the results of the 1998 Lieutenant's Exam to promote Buffalo firefighters to the rank of lieutenant.

As a result of this finding, this court's December 19, 2007 order enjoining the plaintiffs in the New York State Supreme Court action entitled *Margerum, et al. v. City of*

Buffalo, et al., Index No. 1462/2007, and their attorneys, from seeking further relief in the State courts is hereby rescinded, there being no remaining risk of inconsistent judgments or conflicting remedies with respect to the City's promotional practices based on the results of the 1998 Lieutenant's Exam.

A telephone conference with counsel for the parties is scheduled for March 31, 2009, at 10:30 a.m. to discuss a schedule for further proceedings in this action and in the related action, No. 03-CV-580. The court will initiate the call.

So ordered.

\s\ John T. Curtin
JOHN T. CURTIN
United States District Judge

Dated: 3/4/09
p:\opinions\98-99.feb1309